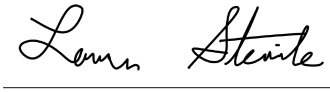
Zachary James

Doctor Shopping and the Patient Sharing Network of Healthcare Providers

Undergraduate Thesis

Nicoleta Serban Ph.D.	Lauren Steimle Ph.D.
	
Signature	*Signature*
04/28/2021	04/28/2021
Date	*Date*

2021

Abstract

Patients with chronic opioid use disorder (OUD) often engage in "doctor shopping" to improperly obtain prescription opioids. An analysis of the patient sharing network of healthcare providers that serve patients with OUD can be used to identify the likelihood that a provider is involved in such behavior. This study relied on Medicaid claims to construct a graph that had healthcare providers as nodes and the number of shared patients as edge weights. It observed the network at the global level using descriptive statistics and at the local level using community detection algorithms. The study then investigated the impact of patient population characteristics on graph variation using exponential random graph models. Statistical analysis revealed a sparse graph where healthcare providers with similar patient demographics were more likely to share an edge. This may indicate that demographics are a key factor in deciding network structure.

Introduction

In 2019, life expectancy in the United States fell for the third straight year.[1] There are multiple causes for this trend, but one of the most concerning is the rise in deaths due to opioid overdoses.[2] Over the past twenty years, prescription opioids have become increasingly commonplace, finding their way into the average American's medicine cabinet. Easy access to the highly addictive substance has led to an increase in opioid use disorder (OUD), and in turn, an epidemic of fatal overdoses.

Most Americans obtain opioids through a prescription written by a physician. Usually, this is to alleviate pain from a medical procedure, such as surgery. However, certain patients schedule appointments and procedures simply to get opioid prescriptions. This practice is known as "doctor shopping".[3] These patients often suffer from OUD and try to obtain opioids to satisfy their addiction.

Researchers have studied "doctor shopping" for several years. However, recent advances in mathematics and computer science have allowed researchers to better understand the problem using network analysis. New models rely on graphical representations of the interactions between patients and physicians. Algorithms then explore the different properties of the graphs. This may reveal that certain types of providers are less likely to prescribe opioids, or that patients in certain regions are more likely to visit multiple physicians to obtain them.

In order to better understand "doctor shopping", a graphical representation of healthcare providers was constructed. Statistical methods were then used to describe the network and determine the relationship between patient population demographics and doctor shopping.

To accomplish this, patients with OUD were matched with the healthcare providers they have visited over the course of a year. This created an undirected weighted graph where two providers share an edge if they share patients, and the weight of the edge is equal to the number of patients shared. Descriptive statistics are used to describe the global nature of the network. Community detection algorithms are used to better understand the local nature of the network. Finally, an exponential random graph model is used to conduct inference and determine the relationship between doctor shopping and patient demographics.

Literature Review

Problems in healthcare have been reduced to graph theory for decades. In 1966, Coleman used social networks to model how doctors come to accept the effectiveness of a new drug.[4] In 1999, researchers in the United Kingdom created a graph of health practitioners to study the diffusion of new medical techniques.[5] Several years later Scott et al. looked at the micro-level, keeping track of every interaction in a clinic over a period of time to create a web of connections.[6] These studies had several common factors. First, the graphs in question were very small, often with fewer than a hundred nodes. Second, the statistical analyses were fairly limited, with the bulk of the analysis consisting of descriptive statistics. These similarities were due to the limitations of graph algorithms at the time.

With advances in computational sciences and graph theory, researchers are now able to conduct more complex analyses on larger graphs. In a 2016 study, researchers examined the effect provider communities have on patients with opioid use disorders.[7] The researchers used a modularity optimization community detection algorithm on a graph of several thousand nodes to determine the provider communities. In a 2019 study, researchers also examined the effect provider communities have on patients. However, the researchers used a multi-scale community detection algorithm, which has been shown to more accurately represent small groups in large data sets.[8]

The application of community detection to healthcare research is relatively new. Meta-analyses have shown that healthcare social analyses are often limited in scope and do not recommend changes to practices.[9] As graph algorithms become more efficient this is sure to change. While this study is not fundamentally different from previous studies, it hopes to expand the current research area by conducting more sophisticated analyses on a relatively large graph.

Data Source

This study relies on the 2012-2013 Medicaid Analytic eXtract (MAX) claims provided by the Centers for Medicare and Medicaid Services (CMS).[10] The data

contains patient-level Medicaid claims for twenty-eight states. The Personal Summary (MAX PS), Inpatient (MAX IP), and Other Therapy (MAX OT) tables were extracted from the data source for analysis. The MAX PS table contains patient characteristics, the MAX IP table provides incident data for patients with inpatient visits, and the MAX OT table provides incident data for patients who have received a service other than an inpatient visit, such as pharmacy services or an outpatient visit.

Healthcare providers are identified in each table with the help of the 2013 National Plan and Provider Enumeration System (NPPES), which assigns each provider a unique National Provider Identifier (NPI). Healthcare providers who were not assigned an NPI by the CMS in the 2013 NPPES are omitted from this study.

The study population consists of Georgia-based healthcare providers who filed Medicaid claims for patients identified with OUD. Several providers in states bordering Georgia (Alabama, Florida, North Carolina, South Carolina, Tennessee) shared OUD patients with Georgia-based providers. These providers have been included in this study. Providers with fewer than eleven OUD patients have been excluded in order to maintain proper anonymization. Data use has been approved by the Georgia Institute of Technology Institutional Review Board (protocol #H11287) and the Centers for Medicare and Medicaid Services (Data Use Agreement #23621).

Methodology

Patients diagnosed with OUD who live in Georgia are identified using the MAX PS table, which lists the diagnosis as a patient attribute. The MAX IP and MAX OT tables are then used to identify the healthcare providers these patients visited in the 2012-2013 year. The healthcare providers are identified by their NPI. For each pair of healthcare providers, the number of shared OUD patients is determined. The result is a table where the first and second columns are an NPI and the third column is the number of patients shared by the two providers. This table is converted into an undirected weighted graph of the form $G = (V, E)$. V is the set of vertices, where each $i \in V$ corresponds to an NPI. E is the set of edges. Each edge is a 3-tuple of the form (i, j, w) where $i, j \in V$ are two distinct vertices and $w \in \mathbb{Z}^+$ is the number of patients shared by the two providers represented by the vertices. Edges where $w < 11$ have been omitted to ensure proper anonymization.

Patient demographics were compiled by aggregating data from the MAX PS table. Demographic information was gathered for four different factors: sex, race, residential urbanicity, and clinical risk group (CRG). Sex and race are determined directly from the MAX PS table. Urbanicity is provided by the United States Department of Agriculture's 2013 Rural-Urban Continuum Codes.[11] Patient CRG levels are calculated using the 3M^{TM} Clinical Risk

Group Methodology, which assigns "each patient to a single, mutually exclusive risk category."[12] Demographic information for certain providers was omitted due to anonymization requirements. During analysis, these omitted values were replaced with the median value of the data.

The graph was constructed and analyzed in Python and R. The graph was analyzed using three different methods: descriptive statistics, community detection, and exponential random graph models (ERGM).

Descriptive Statistics

Descriptive statistics offer insights into the global nature of the network. The density of the graph is defined as the number of edges divided by the maximum number of edges possible:

$$D = \frac{|E|}{\binom{|V|}{2}} = \frac{2|E|}{|V|(|V|-1)}$$

High density indicates that providers are likely to share patients with many other providers. Since the healthcare providers are spread throughout the state of Georgia, a high density may indicate that distance does not prevent providers from sharing patients. Transitivity is defined as the number of triangles in the graph divided by the number of possible triangles.[13] It serves as a measure of whether nodes naturally cluster together. Low transitivity would indicate that the graph as a whole does not cluster well. Finally, the number of components of the graph reveals whether or not the graph is fully connected. If the number of components is greater than one, there are distinct networks within the graph. Descriptive statistical analysis was conducted in R using the *network* package.[14]

Community Detection

Next, the graph is observed at the local level using community detection. A community of a graph $G = (V, E)$ is a subset of V such that any two vertices in the subset are "similar". Analyzing these communities reveals the local structure of the network. Community detection algorithms aim to assign every node to a community. Figure 1 depicts an example graph and three possible communities.[15] This study relied on modularity optimization community detection algorithms. Modularity is a measure of how well a clustering describes the communities in a graph.[16] It is defined as

$$Q = \frac{1}{2m} \sum_{ij} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j)$$

where A_{ij} is the weight of the edge between node i and node j, k_i and k_j are the sum of the weights of i and j's edges, m is the sum of all edge weights in

the graph, c_i and c_j are integers denoting the communities of i and j, and δ is the Kronecker delta function, defined as

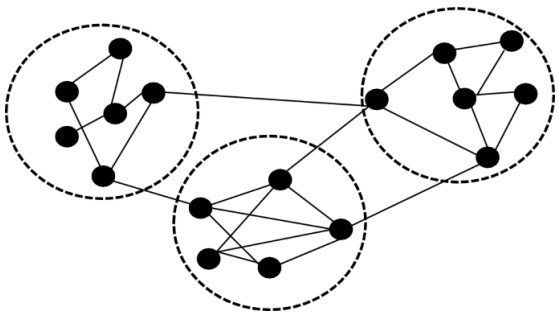
$$\delta(x, y) = \begin{cases} 0 & x \neq y \\ 1 & x = y \end{cases}$$

The resulting value lies in the range $[-\frac{1}{2}, 1]$, where a higher value indicates a partition more representative of the communities in the graph. This leads to the following optimization problem: maximize the modularity according to decision variables $c_1, c_2 \dots c_n$, which correspond to the n vertices in V. In addition, the number of communities must be less than or equal to the number of vertices.

$$\begin{aligned} &\max_{c_1 \dots c_n} \ Q \\ &\text{s.t. } c_i \in \mathbb{Z} && i = 1 \dots n \\ &\quad 1 \leq c_i \leq n && i = 1 \dots n \end{aligned}$$

Calculating the modularity for every single possible partition is computationally infeasible. Therefore, the algorithms used in this study rely on heuristics to efficiently find a partition with high modularity. After determining the communities within the graph, statistical descriptors of each subgraph are analyzed. Community detection was performed in Python using the *igraph* package.[17]

Figure 1: Example of community detection.



Exponential Random Graph Model

Finally, the relationship between the graphical characteristics of a node and the characteristics of its patient population are explored. With other forms of data, a simple regression can be used to determine the relationship. However, the nodes in a graph are neither independent from one another nor identically distributed. Therefore, a different model is used. Below is a model from the exponential family.

$$P(Y = y \mid \theta) = \frac{\exp(\theta^T s(y))}{c(\theta)}, \quad \forall y \ \in Y$$

The right side is the probability of observing graph y from random variable Y given parameters θ. This probability is modeled with the function on the right, where $s(y)$ are attributes of the derived graph and $c(\theta)$ is a normalization factor. Using the model, inference was conducted on the network. In particular, attributes of the healthcare providers' patient population were included in $s(y)$. The factors are

- Percentage of population that is female
- Percentage of population that is white
- Percentage of population that lives in a rural area
- Percentage of population that is in a low or medium clinical risk group (CRG)

For each factor, we determine the relationship between the difference of two node's values and the probability they share an edge. The null hypothesis is that the probability of there being an edge between nodes v and u is the same if they have the minimum difference in factor value ($u_{factor} - v_{factor} = 0$) and if they have the maximum difference in factor value ($u_{factor} - v_{factor} = 1$). The alternative hypothesis is that these probabilities are unequal. ERGM analysis was conducted in R using the *statnet* package.[18]

Results

Network Characteristics

The resulting graph contained $n = 22,062$ different providers. These providers were connected to one another by $1,151,912$ edges. The edges have a median weight of 31, which has been inflated by the omission of edges with weight less than eleven. The graph is also connected, containing one component and no isolates. This may also be the result of data anonymization. Providers with a small number of patients are more likely to have a small degree and be part of an isolated subgraph. However, all providers with fewer than eleven patients have been removed from the data set. Furthermore, several providers have very high degrees. These outliers prevent the formation of isolated subgraphs.

Global Structure

The resulting graph was very sparse, with density $D = 0.00473$. This can be explained by the nature of patient-provider interactions. Patients must physically visit a provider to obtain an opioid prescription. As a result, a patient is more likely to obtain prescriptions by visiting a few nearby providers as opposed to many far-away providers. This is especially the case in Georgia, which is a geographically large state. The graph also had very low transitivity, with $T = 0.10074$. This indicates that the nodes are not closely clustered. More concretely, if provider u shares patients with provider v, and provider v shares

patients with provider w, the probability of provider u sharing patients with provider w is relatively low. There are many possible explanations for this. For example, a patient at provider u may simply be unable to visit provider w due to geographic constraints.

Local Structure

Several different modularity optimization community detection algorithms were run on the graph. The modularities of the resulting partitions are reported in Table 1. The partition returned from the Louvain algorithm had the highest modularity and was used for analysis. The high modularity returned by the algorithm indicates that the model was a good fit. The algorithm identified fourteen communities. Characteristics of the ten largest communities are reported in Table 2. The patient demographics of each community do not vary significantly from the overall patient demographics. This indicates that another factor may be responsible for the formation of distinct communities. For example, patients in certain regions of the state may be more likely to visit a specific subset of providers.

Table 1: Community Detection Results

Algorithm	Greedy	Neumann	Louvain
Modularity	0.10518	0.33869	0.52058

Table 2: Composition of Ten Largest Communities

Group	Community Size	Density
1	3976	0.015921
2	3069	0.023825
3	2104	0.033251
4	1032	0.057452
5	1211	0.061116
6	1596	0.037905
7	1807	0.031887
8	1437	0.040777
9	1886	0.046189
10	1607	0.014660

Exponential Random Graph Model

Table 3 contains the results of the ERGM predicting the log-odds of two providers with minimum factor value difference being connected by an edge. The analysis revealed high homophily among all factors ($p < 0.001$). This means that providers with similar patient demographics were more likely to share patients. This result was not unexpected for most of the factors.

Sex

Providers often cater to patients of one particular sex, so it is reasonable to assume that providers with similar patient sex ratios are more likely to share patients. For example, an oncologist who specializes in breast cancer may share patients with a gynecologist.

Race

At a high level, the state of Georgia can be divided into two regions: the Atlanta metropolitan area, which is predominantly African American, and the rural counties in the remainder of the state, which are predominately white. This geographic difference in demographics explains why race can indicate a higher probability of patient sharing.

Urbanicity

Patients living in rural areas constituted a relatively small portion of the total number of patients. The presence of such patients in a provider's patient population indicates that the provider itself may be located in a rural area. This increases the probability that the provider shares patients with other rural providers.

Clinical Risk Groups

Providers with similar proportions of patients in low or medium clinical risk groups were more likely to share patients. It is not clear why this is the case. Providers may be more likely to treat patients in a certain CRG, similar to the explanation for the female percentage factor. Additional research will be needed.

Table 3: ERGM Predictions

Factor	Coefficient	Standard error
Sex	1.956139***	0.002516
Race	0.653009***	0.002697
Urbanicity	0.557561***	0.003428
Clinical Risk Group	-0.232271***	0.00521

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Discussion

The above analysis provides several insights into the patient sharing networks of healthcare providers who serve chronic opioid users. First, it reveals that such networks are highly sparse, and do not cluster naturally. Second, it reveals that they can be partitioned into distinct communities that are themselves highly connected with one another. Finally, it shows that the demographics of a patient population directly affect the degree to which a provider is involved in doctor shopping.

These results can inform public health interventions. Given a patient diagnosed with OUD, public health officials may wish to determine where the patient obtained their prescriptions. Considering the sparse nature of the patient sharing network, the patient most likely visited only a handful of providers. Furthermore, there is reason to believe that all the providers would be located in the same geographic area. This claim should be further investigated with social network analysis that incorporates provider locations.

Healthcare providers in the patient sharing network can be partitioned into distinct communities. These communities are significantly more dense than the graph as a whole. Additionally, the groups themselves are highly interconnected. Further research is needed to determine why these groups form. As above, geographic location may be the determining factor. Providers within a certain geographic area may have a greater probability of being in the same community. For example, a community may consist of providers from the Atlanta metropolitan area, or northwest Georgia, centered around Dalton. It would then be reasonable for there to be patients who visit these two groups, explaining why communities are connected.

Finally, having taken into account the previous two results, public health officials may wish to predict the degree to which a provider is involved in doctor shopping. More precisely, given two providers and their patient demographics, public health officials may wish to know the probability that the two providers share patients with OUD. This can be accomplished with the ERGM. The model revealed that providers with similar patient demographics are more likely to share patients. Furthermore, given the demographics of two providers, the log odds of the two providers sharing patients can be calculated.

These statistical results give public health officials new tools in the fight against opioid addiction. They enable a better understanding of doctor shopping, one of the leading causes of opioid abuse. Future research could incorporate detailed electronic health records. This would enable normal patient behavior to be differentiated from doctor shopping. In addition, a more detailed graph could be created by including individual patients as nodes instead of aggregating by provider. While further research is needed, the above results will hopefully provide additional insight into the causes of the opioid epidemic.

References

1. U.S. life expectancy declining due to more deaths in middle age. Reuters. https://www.reuters.com/article/us-health-life-expectancy/u-s-life-expectancy-declining-due-to-more-deaths-in-middle-age-idUSKBN1Y02C7. Published November 26, 2019. Accessed October 21, 2020.

2. Gold MS. The Role of Alcohol, Drugs, and Deaths of Despair in the U.S.'s Falling Life Expectancy. Mo. Med. 2020 117(2): 99–101. https://pubmed.ncbi.nlm.nih.gov/32308224/. Published March 2020. Accessed November 18, 2020

3. Doctor Shopping Laws. CDC. https://www.cdc.gov/phlp/docs/menu-shoppinglaws.pdf. Published 28, 2012. Accessed November 8, 2020.

4. Coleman JS, Katz E, Menzel H. Medical innovation: A diffusion study. Behav. Sci. Rep. 1966 https://doi.org/10.1002/bs.3830120608

5. West Elizabeth, Barron DN, Dowsett J. Hierarchies and cliques in the social networks of health care professionals: implications for the design of dissemination strategies. Soc. Sci. Med. 1999;48(5):633-646 https://doi.org/10.1016/S0277-9536(98)00361-X

6. Scott J, Tallia A, Crosson JC, Orzano AJ, Stroebel C, DiCicco-Bloom B, O'Malley D, Shaw E, Crabtree B. Social Network Analysis as an Analytic Tool for Interaction Patterns in Primary Care Practices. Annals of Family Medicine 2005;3(5): 443-448. https://doi.org/10.1370/afm.344

7. Stein BD, Mendelsohn J, Gordon AJ, Dick AW, Burns AW, Sorbero M, Shih RA, Pacula RL. Opioid analgesic and benzodiazepine prescribing among Medicaid-enrollees with opioid use disorders: The influence of provider communities. Journal of Addictive Diseases. 2017; 36(1): 14-22 https://doi.org/10.1080/10550887.2016.1211784

8. Ostovari M, Yu D. Impact of care provider network characteristics on patient outcomes: Usage of social network analysis and a multi-scale community detection. PLoS ONE. 2019 https://doi.org/10.1371/journal.pone.0222016

9. Chambers D, Wilson P, Thompson C, Harden M. Social Network Analysis in Healthcare Settings: A Systematic Scoping Review. PLoS ONE. 2012;7(8). https://doi.org/10.1371/journal.pone.0041911

10. https://www.cms.gov/Research-Statistics-Data-and-Systems/Computer-Data-and-Systems/MedicaidDataSourcesGenInfo/MAXGeneralInformation

11. https://www.ers.usda.gov/data-products/rural-urban-continuum-codes.aspx

12. https://www.3m.com/3M/en_US/health-information-systems-us/drive-value-based-care/patient-classification-methodologies/crgs/

13. Wasserman S, Faust K. Social Network Analysis: Methods and Applications. Cambridge: Cambridge University Press; 1994.

14. Butts C. A Package for Managing Relational Data in R. Journal of Statistical Software. 2008;24(2). http://dx.doi.org/10.18637/jss.v024.i02

15. Al-Andoli M, Cheah WP, Tan SC. Deep learning-based community detection in complex networks with network partitioning and reduction of trainable parameters. J Ambient Intell Human Computing 2021; 12, 2527–2545. https://doi.org/10.1007/s12652-020-02389-x

16. Newman ME. Fast algorithm for detecting community structure in networks. Proc. Natl. Acad. Sci. USA 2004;69(6) https://doi.org/10.1103/PhysRevE.69.066133

17. Csardi G, Nepusz T. The igraph software package for complex network research. InterJournal, Complex Systems 2006;1695. https://igraph.org

18. Hunter D, Handcock M, Butts C, Goodreau S, Morris M. ergm: A Package to Fit, Simulate and Diagnose Exponential-Family Models for Networks. Journal of Statistical Software. 2008;24(3),1-29. https://doi.org/10.18637/jss.v024.i03